IJARCCE

ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 12, December 2021

DOI: 10.17148/IJARCCE.2021.101237

SURVEY ON VIDEO CONFERENCING,HAND GESTURE RECOGNITION AND AIR WRITING

Deepak K.N.[1], Anand Ramesh[2], Manu T.M.[3], Suraj K.S.[4], Vaishnavu M.V.[5]

Assistant Professor, Department of Computer Science And Engineering,Universal Engineering College ,Vallivattom, Thrissur, India.[1]

B.Tech Student, Department of Computer Science And Engineering,Universal Engineering College ,Vallivattom, Thrissur, India.[2,3,4,5]

Abstract:Nowadays, due to pandemic, many schools or colleges have switched to remote class via video conferencing. For this situation, a video conference with higher features are to be implemented.In this survey the approach to improve video conference are discussed here.

Keywords: Video Conference, Hand gesture recognition, Air-writing

I. INTRODUCTION

The main goal of this survey is to make video conference more interactive and easy to use.Physical interaction between humans has steadily decreased over the past year as a result of the pandemic. As a result, the adoption of video conference has seen a huge increase in conducting business globally and technologically. Using video conference in learning meets the primary goal of both educators and students. Through web conferencing, higher learning institutions have the ability to capitalize on the available technologies to expand access to instructors online while also creating new experiences in the teaching and learning environment. With the rise in demand of the video conference it is necessary to be improved further.

II. THEORY

A. Video Conference

Video conference is a type of online meeting where two or more people engage in a live audio-visual call. With a strong internet connection, the participants can see, hear, and talk to each other in real time, no matter where in the world they are.In business, people typically use video conferencing to communicate and collaborate within and outside an organization. It only needs the necessary hardware and software set up to get the most out of the experience.

B. Hand Gesture Recognition

Hand Gesture recognition is an alternative user interface for providing real-time data to a computer. Instead of typing with keys or tapping on a touch screen, a motion sensor perceives and interprets movements as the primary source of data input. This is what happens between the time a gesture is made and the computer reacts.A camera feeds image data into a computer.A Specially designed software identifies meaningful gestures from a predetermined gesture library where each gesture is matched to a computer command.The software then correlates each registered real-time gesture, interprets the gesture and uses the library to identify meaningful gestures that match the library.Once the gesture has been interpreted, the computer executes the command correlated to that specific gesture.

C. Air writing

Air writing is defined aswriting alphanumeric with hand or finger movements in athree-dimensional (3D) free space. It is particularlyuseful for user interfaces that do not allow the user to type on the keyboard or write on the touchpad/touch screen or for textinput for intelligent system control.Air-writing recognition falls into three categories from the perspective of data acquisition: vision-based, sensor-based, and WiFi signals based. Vision based schemes use cameras to capture texts written by human users. Sensor based schemes use hand-held sensors, such as Wiimote and

Leap Motionto capture user hand movements. WiFi signal based schemes use WiFi signals to recognize user handwritings.

III. RELATED WORK

Here we introduce each papers based on the technologies used in the AI-Enabled Video Conference and they are arranged in technologies bases
In this paper [1], a P2P-MCU approach is proposed for multi-party video conferencing that efficiently supports both ordinary smart mobile phones and PCs. By this approach, a MCU module is integrated into the browser to mix and transcode the video & audio streams in real time. And when the browser acts as the MCU, the node leaves the conference session without notice, another candidate browser can take over the control immediately, and the ongoing WebRTC conference can be seamlessly recovered with an MCU selection algorithm. In addition to this, the proposed system works under the 3G symmetric NAT networks by using some UDP hole punching method. This P2P-MCU solution reduces 64% CPU usages and 35% bandwidth consumptions for each participant compared to the mesh-network solution in an eight-party WebRTC conference experiments. Although the P2P-MCU module may introduce some delay (<500ms), the delay is stable and perceptually almost neglect able.
According to this paper [2], theWebRTC protocol is restricted to a small number of peers because there is no simple way to mix real-time streams from multiple peers and then distribute the mixed stream to a large number of participants. For example, it is necessary to mix audio and video streams from peers in a conversation and broadcast the real-time mixed stream to more than10k participants in the conference. This paper proposes a method for the synchronized mixing of real-time audio/video streams from multiple peers while minimizing latency.Compared to the previous WebRTC architecture, an additional quasi-peer is added to the WebRTC gateway. The quasi-peer functions as a normal peer in that it collects media streams from other peers, but it does not generate any media streams for other peers. After the audio/video data are collected, the quasi-peer mixes the data and ultimately distributes the output(the final mixed audio/video)via ICE, a content delivery network to the participants.This method enables the implementation of an online live conversation system that is able to mix live conversation streams from multiple peers and then rebroadcast the mixed stream to a large number of participants.
In an enterprise network, it is common to have hundreds of video conferences held simultaneously such that a large number of video streams need to be transmitted between participants in different geographic allocations.For such a large transmission demand,an advance reservation(AR)system deployed for the video conferencing system can provide quality-of-service (QoS) guarantees to users and improve the resource utilization of the network.
In this paper[3], the authors propose an algorithm called the elastic time slot-based advance reservation algorithm (ETARA), which aims at improving the resource utilization and reducing the computational complexity. The Advanced Reservation(AR) algorithm makes use of the topology management module (TMM) to obtain the topology of the network and the traffic engineering database(TED)to obtain the available bandwidth on each link. Combined with the request time, the AR algorithm generates resource matrices with the time attribute. As the request is processed, the AR algorithm makes a decision based on whether there issufficient bandwidth in resource matrices to ensure a certain QoS level.The results show that with the same acceptance ratio, the runtime of ETARA can be up to 57 times lower than that of the flexible time slot-based approach.Though ETARA has a slightly longer run time than the dynamic approach, the acceptance ratio of ETARA can be twice as high as that of the dynamic timeslot-based approach.
In this paper[4], the authors has proposed a method for maximizing the number of admitted Dynamic Multicast requests in the Advance Reservation environment (MDMAR) for the enterprise video conferencing system.The abilities to reserve resources in advance, as well as effective dynamic multicast when participants can join and leave the conference at any time, are essential in the distributed multiparty video conferencing systems. However, the effective advance reservation strategies of the dynamic multicast requests for a heavy traffic case still remains open. For this two path schemes of a fixed path and variable paths, as well as a heterogeneous bandwidth reservation model is taken into account. The MDMAR problem is NP-complete and formulate it mathematically as an integer linear program (ILP) for small networks. Then, greedy algorithms and simulated annealing (SA) algorithms for enterprise networks is developed. Comparative simulations are performed to evaluate the heuristic algorithms for both small networks and enterprise networks. From that it is found that the SA algorithms can provide within 6% lower optimal solutions than the ILP algorithms for small network, and up to 10% improvement over the greedy algorithms for the large campus or enterprise network.
The main aim of this paper [5] is to maximize the received video quality for both systems under uplink-downlink capacity constraints, while constraining the number of hops the packets traversetotwo.One way to deal with user bandwidth heterogeneity is employing layered video coding, generating multiple layers with different rates, whereas an alternative is partitioning the receivers of each source and disseminating a different non-layered video version withineachgroup. Here authors has proposed analgorithmthatsolves for the number of video layers, layer rates, and distribution trees for the layered system. For the partitioned simulcast system, an algorithm is developed to determine the receiver partitions along with the video rate and the distribution trees for each group. Through numerical

comparison, we show that the partitioned simulcast system achieves the same average receiving quality as the ideal layered system without any coding overhead for the four-user systems simulated, and better quality than the layered system when the layered coding overhead is only 20%. The two systems perform similarly for the six-user case if the layered coding overhead is 10%.

According to the paper [6], increasingly end-hosts in a multi-user video conference are assisted by cloud-based servers that improve the quality of experience for end users. For this, a proposed system is introduced to evaluate the impact of strategies for placement of such servers on user experience and deployment cost. The authors consider scenarios based upon the Amazon EC2 infrastructure as well as future scenarios in which cloud instances can be located at a larger number of possible sites across the planet. The proposed system is driven by real data to create demand scenarios with realistic geographical user distributions and diurnal behaviour. According to the proposed system it is found that on the EC2 infrastructure a well-chosen static selection of servers performs well but as more cloud locations are available a dynamic choice of servers becomes important.

This paper[7] proposes methods for rate adaptation by motion-based spatial and temporal resolution selection in both mesh-connected and selective-forwarding-unit (SFU) connected WebRTC videoconferencing using scalable video coding. In the mesh-connected case, the proposed motion-adaptive spatial/temporal layer selection allows each peer to send video to different peers with different terminal types and network rates at different rates using a single encoder. In the SFU-connected case, motion-adaptive rate control is used both at peers to adapt to the network rate between the sending peer and SFU by spatiotemporal resolution adaptation and at the SFU by layer selection to adapt to the network rate between the SFU and receiving peer. Experimental results show that our proposed motion-based rate adaptation achieves better perceptual video quality with sufficiently high frame rates and lower quantization parameter for video with high motion; and high spatial resolution and lower quantization parameter for video with low motion compared to simple rate-distortion model based layer selection that does not use motion complexity, at the same rate.

In the paper [8] a device consists of 40 microphones to be worn at the wrist is introduced. The gesture recognition performance is evaluated through the identification of 36 gestures in American sign language (ASL), including 26 ASL alphabetical characters and 10 ASL numbers. The optimal area for sensor band placement (distal/proximal) is examined to reveal the location of the highest discrimination accuracy. Ten subjects are recruited to perform over ten trials for each set of hand gestures. The results of the paper shows that the intra subject average classification accuracy above 90% using the two features with all 40 microphones, while the average classification accuracy exceeding 84% is obtained using ten microphones. These results indicate that acoustic signatures from the human wrist can be utilized for hand gesture recognition, while the use of few, simple features, with low computational requirements is sufficient to characterize some hand gesture.

This paper [9] introduces a hand gesture recognition sensor using ultra-wideband impulse signals which are reflected from a hand. The reflected waveforms in time domain are determined by the reflection surface of a target. Thus every gesture has its own reflected waveform. Hence the proposed systemuses machine learning approach such as convolutional neural network (CNN) for the gesture classification. The CNN extracts its own feature and constructs classification model then classifies the reflected waveforms. Six hand gestures from American Sign Language (ASL) are used for an experiment and the result shows more than 90% recognition accuracy. For fine movements, a rotating plaster model is measured with 10° step. An average recognition accuracy is also above 90%.

In this paper [10] a prototype system, including a wearable gesture sensing device with four pressure sensors and the corresponding algorithmic framework, is developed to realize real-time gesturebased interaction. With the device worn on the wrist, the user can interact with the computer using 8 predefined gestures. Experimental results show that the delay of gesture recognition is about 100ms, with the average accuracy of 95.28% in the experienced-user test and 86.20% in the inexperienced-user test. Finally the system is evaluated by a mouse-controlling interaction task and performs well. Both experienced and inexperienced people can easily and quickly complete interactive tasks. These results demonstrate that a pressure-sensor based wristband can be used to classify hand gestures well and to control the mouse interaction. This approach provides an interactive way to replace the mouse for decreasing the risk of the carpal tunnel syndrome (CTS).

In this paper [11], a multimodal hand gesture detection and recognition system using differential Pyroelectric Infrared (PIR) sensors and a regular camera is described. Any movement within the viewing range of the differential PIR sensors are first detected by the sensors and then checked if it is due to a hand gesture or not by video analysis. If the movement is due to a hand, one-dimensional continuous-time signals extracted from the PIR sensors are used to classify/recognize the hand movements in real-time. Classification of different hand gestures by using the differential PIR sensors is carried out by a new winner-takeall (WTA) hash based recognition method. Jaccard distance is used to compare the WTA hash codes extracted from 1-D differential infrared sensor signals. It is experimentally shown that the multimodal system achieves higher recognition rates than the system based on only the on/off decisions of the analog circuitry of the PIR sensors.

In this paper[12], a novel system is proposed for dynamic hand gesture recognition using multiple deep learning architectures for hand segmentation, local and global feature representations, and sequence feature globalization and recognition. The proposed system is evaluated on a very challenging dataset, which consists of 40 dynamic hand gestures performed by 40 subjects in an uncontrolled environment. The results show that the proposed system outperforms state-of-the-art approaches, demonstrating its effectiveness.

In this paper[13],we propose a micro hand gesture recognition system and methods using ultrasonic active sensing. This system uses micro dynamic hand gestures for recognition to achieve human–computer interaction (HCI). The implemented system, called hand-ultrasonic gesture (HUG), consists of ultrasonic active sensing, pulsed radar signal processing, and time-sequence pattern recognition by machine learning. A lower frequency(300kHz)ultrasonic activesensing to obtain high resolution range-Doppler image features is adopted here. Using high quality sequential range-Doppler features, a state-transition-based hidden Markov model is proposed for gesture classification. This method achieves a recognition accuracy of nearly 90% by using symbolized range-Doppler features and significantly reduces the computational complexity and power consumption.Further more,to achieve higher classification accuracy,we utilize an end-to-end neural network model and obtain a recognition accuracy of 96.32%.

In this paper [14], a real-time dynamic finger gesture recognition using a soft sensor embedded data glove is presented, which measures the metacarpophalangeal (MCP) and proximal interphalangeal (PIP) joint angles of five fingers. In the gesture recognition field, a challenging problem is that of separating meaningful dynamic gestures from a continuous data stream. To solve the problem of separating meaningful dynamic gestures, the authors has proposed a deep learning-based gesture spotting algorithm that detects the start/end of a gesture sequence in a continuous data stream. The gestures potting algorithm takes window data and estimates a scalar value named gesture progress sequence (GPS). Moreover, to solve the gesture variation problem, a sequence simplification algorithm and a deep learning-based gesture recognition algorithm is proposed here. The proposed three algorithms (gesture spotting algorithm, sequence simplification algorithm, and gesture recognition algorithm) are unified into the real-time gesture recognition system and the system was tested with 11 dynamic finger gestures in real-time. The proposed system took only 6 ms to estimate a GPS and no more than 12 ms to recognize the completed gesture in real-time.

In this paper[15], a frame work,using deep-learning techniques, is proposed here toclassifyhand-gesture signatures generated from an ultra-wideband (UWB) impulse radar. The signals of 14 different hand-gestures are extracted and represent each signature as a 3-dimensional tensor consisting of range-Doppler frame sequence. These signatures are passed to a convolutional neural network (CNN) to extract the unique features of each gesture, and are then fed to a classifier. Four different classification architectures to predict the gesture class, namely; fully connected neural network (FCNN), k-Nearest Neighbours (k-NN), support vector machine (SVM), (iv) long short term memory (LSTM) network is compared here. The shape of the range-Doppler-frame tensor and the parameters of the classifiers are optimized in order to maximize the classification accuracy. The classification results of the proposed architectures show a highlevel of accuracy above 96% and a very low confusion probability even between similar gestures.

In this paper[16], the CSI (channel state information)derived from wireless signals to realize the device-free air-write recognition called WriFi is utilized for air writing purpose. Compared to the gesture recognition, the increased diversity and complexity of characters of the alphabet make it challenging. The PCA (Principle Component Analysis) is used for denoising effectively and the energy indicator derived from the FFT (Fast Fourier Transform) is to detect action continuously. The unique CSI waveform caused by unique writing patterns of 26 letters serve as feature space. From the experiments conducted in the laboratory the average accuracy of the Wri-Fi are 86.75% and 88.74% in two writing areas, respectively.

In this paper[17], an air-writing system using acoustic waves is proposed. The proposed system consists of two components: a motion tracking component, and a text recognition component. For motion tracking, we utilize direction-of-arrival (DoA) information. An ultrasonic receiver array tracks the motion of a wearable ultrasonic transmitter by observing the change in the DoA of the signals. A novel 2-D DoA estimation algorithm is proposed that can track the change in the direction of the transmitter using measured phase-differences between the receiver array elements. The proposed phase-difference projection (PDP) algorithm can provide accurate tracking with a 3-sensor receiver array. The motion tracking information is passed next for text recognition. To this end, and in order to strike the desired balance between flexibility, processing speed, and accuracy, a training-free order restricted matching (ORM) classifier is designed. The proposed air-writing system, which combines the proposed DoA estimation and text recognition algorithms, achieves a letter classification accuracy of 96.7%. The utility, processing time, and classification accuracy are compared with four training-free classifiers and two machine learning classifiers to demonstrate the efficiency of the proposed system.

This work of the paper [18], presents a prototype framework for vision-based mid-air unistroke character input, which can be adapted as an interface for the IRS. At first, an acquisition module is developed which effectively spots the legitimate gesture trajectory by implementing pen-up and pen-down actions using depth thresholding and velocity tracking. The extracted trajectory is recognized through a novel, fast, and easy to implement the equipolar signature (EPS) technique. Apart from resistance to rotation, scale, and translation variations, EPS exhibits neutrality to stroking directions as well. On the three self-collected datasets comprising of digits, alphabets, and symbols, the EPS scheme obtains over 96.5% accurate results with an average of 30-ms running time. The proposed scheme is also validated on an open dataset DAIR (Dataset for AIR Handwriting), where it achieves 95.5% mean accuracy with 24.3-ms recognition time per gesture. The developed approach is compared with benchmark schemes to justify its accuracy and speed.

In this paper[19], recognition of characters or words is accomplished based on six-degree-of freedom hand motion data. Air-writing are of two levels: motion characters and motion words. Isolated air-writing characters can be recognized similar to motion gestures although with increased sophistication and variability. For motion word recognition in which letters are connected and superimposed in the same virtual box in space, statistical models are built for words by concatenating clustered ligature models and individual letter models. A hidden Markov model is used for air-writing modelling and recognition. The proposed system achieves a word error rate of 0.8% for word-based recognition and 1.9% for letter-based recognition.

This paper[20] addresses detecting and recognizing air-writing activities that are embedded in a continuous motion trajectory without delimitation. Detection of intended writing activities among superfluous finger movements unrelated to letters or words presents a challenge that needs to be treated separately from the traditional problem of pattern recognition. At first a dataset that contains a mixture of writing and non-writing finger motions is made. The LEAP from Leap Motion is used for marker-free and glove-free finger tracking. A window-based approach is proposed that automatically detects and extracts the air-writing event in a continuous stream of motion data, containing stray finger movements unrelated to writing. Consecutive writing events are converted into a writing segment. The recognition performance is further evaluated based on the detected writing segment. The proposed system achieves an overall segment error rate of 1.15% for word-based recognition and 9.84% for letter-based recognition.

CONCLUSION

In summary, we conclude that Video conferencing is one of the best ways of communication for large organizations as they provide an instant and reliable method through which the entire organization can connect, communicate, and collaborate. The Video conferencing holds great promise for new inventions n communication and it can furthe extend its capabilties. Also hand gesture recognition and air writing are most important and challenging steps to a new human computer interaction.

REFERENCE

[1]. Kwok-Fai Ng, Man -Yan Ching, Yang Liu, Tao Cai, Li Li, and Wu Chou, “A P2P-MCU Approach to Multi-Party Video Conference with WebRTC,” International Journal of Future Computer and Communication, Vol. 3, No. 5, October 2014.

[2]. Dongming Tang, and Liqun Zhang “Audio and Video Mixing Method to Enhance WebRTC,” IEEE Access, April. 2020.

[3]. Zhiwen Liao, and Ling Zhang “Elastic Timeslot-based Advance Reservation Algorithm for Enterprise Video Conferencing Systems,” IEEE Access, vol. 4,pp 1-17 ,2019.

[4]. Zhiwen Liao, and Ling, “Scheduling Dynamic Multicast Requests in Advance Reservation Environment for Enterprise Video Conferencing Systems,” IEEE Access, vol. 4, 2016.

[5]. EymenKurdoglu, Yong Liu, and Yao Wang, “DealingWithUserHeterogeneityinP2PMulti-Party Video Conferencing: Layered Distribution Versus Partitioned Simulcast,” IEEETransactions on Multimedia, vol.18, No.1,pp 90-101, Jan 2016.

[6]. Richard G. Clegg, Raul Landa, David Griffin, Miguel Rio, Member, Peter Hughes, Ian Kegel, Tim Stevens, Peter Pietzuch, and Doug Williams, “Faces in the Clouds: Long-Duration, Multi-User, Cloud-Assisted Video Conferencing,” IEEE Transactions on Cloud Computing, 2016,DOI 10.1109/TCC.2017.2680440

[7]. Gonca Bakar, RizaArdaKirmizioglu, and A. Murat Tekalp, “Motion-Based Rate Adaptation in WebRTC Videoconferencing using Scalable Video Coding,” IEEE Transactions on Multimedia, DOI 10.1109/TMM.2018.285662

[8]. Nabeel Siddiqui, and Rosa H. M. Chan, “Hand Gesture Recognition Using Multiple Acoustic Measurements at Wrist,” IEEE Transactions on human-machine systems, vol. 51, No. 1,pp 56-62, Feb. 2021.

[9]. Seo Yul Kim, Hong Gul Han, Jin Woo Kim, Sanghoon Lee, and Tae Wook Kim, “A Hand Gesture Recognition Sensor Using Reflected Impulses,” IEEE Sensors Journal, DOI 10.1109/JSEN.2017.2679220.

[10]. YuFei Zhang, Bin Liu, and Zhiqiang Liu, “Recognizing Hand Gestures with Pressure Sensor based Motion Sensing,” IEEE Transactions on Biomedical Circuits and Systems 1, DOI 10.1109/TBCAS.2019.2940030.

[11]. FatihErden and A. EnisÇetin, Fellow, “Hand Gesture Based Remote Control System Using Infrared Sensors and a Camera,” IEEE Transactions on Consumer Electronics, Vol. 60, No. 4, November 2014,pp 675-680.

[12]. Munneer Al-Hammadi, Ghulam Muhammad, Wadood Abdul, Mansour Alsulaimani, Mohammed A. Bencherif, Tareq S. Alrayes, Hassan Mathkour, and Mohamed Amine Mekhtiche “Deep Learning-Based Approach for Sign Language Gesture Recognition With Efficient Hand Gesture Representation,” IEEE Access, vol. 8,Nov. 2020.

[13]. Yu Sang , Laixi Shi , and Yimin Liu, “Micro Hand Gesture Recognition System Using Ultrasonic Active Sensing,” IEEE Access, vol. 6,3-28 Sept 2018

[14]. Minhyuk Lee, and Joonbum Bae, “Deep Learning Based Real-Time Recognition of Dynamic Finger Gestures Using a Data Glove,” IEEE Access, vol. 8, 17-19 Nov , 2020.

[15]. SruthySkaria, Akram Al-Hourani , and Robin J. Evans, “Deep-Learning Methods for Hand-Gesture Recognition Using Ultra-Wideband Radar,” IEEE Access, vol. 8, 10-19 Nov, 2020.

[16]. Zhangjie Fu, Jiashuang Xu, Zhuangdi Zhu, Alex X. Liu and Xingming Sun, “Writing in the Air with WiFi Signals for Virtual Reality Devices,” IEEE Transactions on Mobile Computing, DOI 10.1109/TMC.2018.2831709.

[17]. Hui Chen, TarigBallal, Ali H. Muqaibel, Xiangliang Zhang, and Tareq Y. Al-Naffouri, “Air-writing via Receiver Array Based Ultrasonic Source Localization,” IEEE Transactions on Instrumentation and Measurement,2020, DOI 10.1109/TIM.2020.299157.

[18]. Lalit Kane and Pritee Khanna, “Vision-Based Mid-Air Unistroke Character Input Using Polar Signatures,” IEEE Transactions on Human-Machine Systems, 2017. DOI 10.1109/THMS.2017.270669.

[19]. Mingyu Chen, GhassanAlRegib, andBiing-Hwang Juang, “Air-Writing Recognition—Part I: Modeling and Recognition of Characters, Words, and Connecting Motions,” IEEE Transactions on Human-Machine Systems,2015, DOI 10.1109/THMS.2015.2492598.

[20]. Mingyu Chen, GhassanAlRegib, and Biing-Hwang Juang, “Air-Writing Recognition—Part II: Detection and Recognition of Writing Activity in Continuous Stream of Motion Data,” IEEE Transactions on Human-Machine Systems, 2015, DOI 10.1109/THMS.2015.2492599.